

Top Ten Seven Reasons we use SMW at PNNL

Keith Fligg

Pacific Northwest National Laboratory

keith.fligg@pnl.gov



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

#7: Knowledge Encapsulation Framework.

- ▶ KEF – Knowledge Encapsulation Framework
 - Lab Directed Research and Development
- ▶ Features
 - SMW + Blog (Wordpress) + Forum (PHPBB)
 - Single Sign On via Kerberos
- ▶ Growth
 - About 15-20 sites using KEF
 - Growing pains moving from research task to client-ready
 - Kerberos (for SSO) was slow
 - Needed a build system



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

#6: A build system.

- ▶ Jenkins
- ▶ Ant
- ▶ Subversion
 - core – Core MediaWiki, Wordpress and PHPBB
 - core_extensions – SMW, Halo, SemanticForms, etc.
 - pnnl – PNNL global code and changes to above
 - <site> – Site specific code and changes to above
- ▶ Setup scripts
- ▶ KEF 'Modules'



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Jenkins

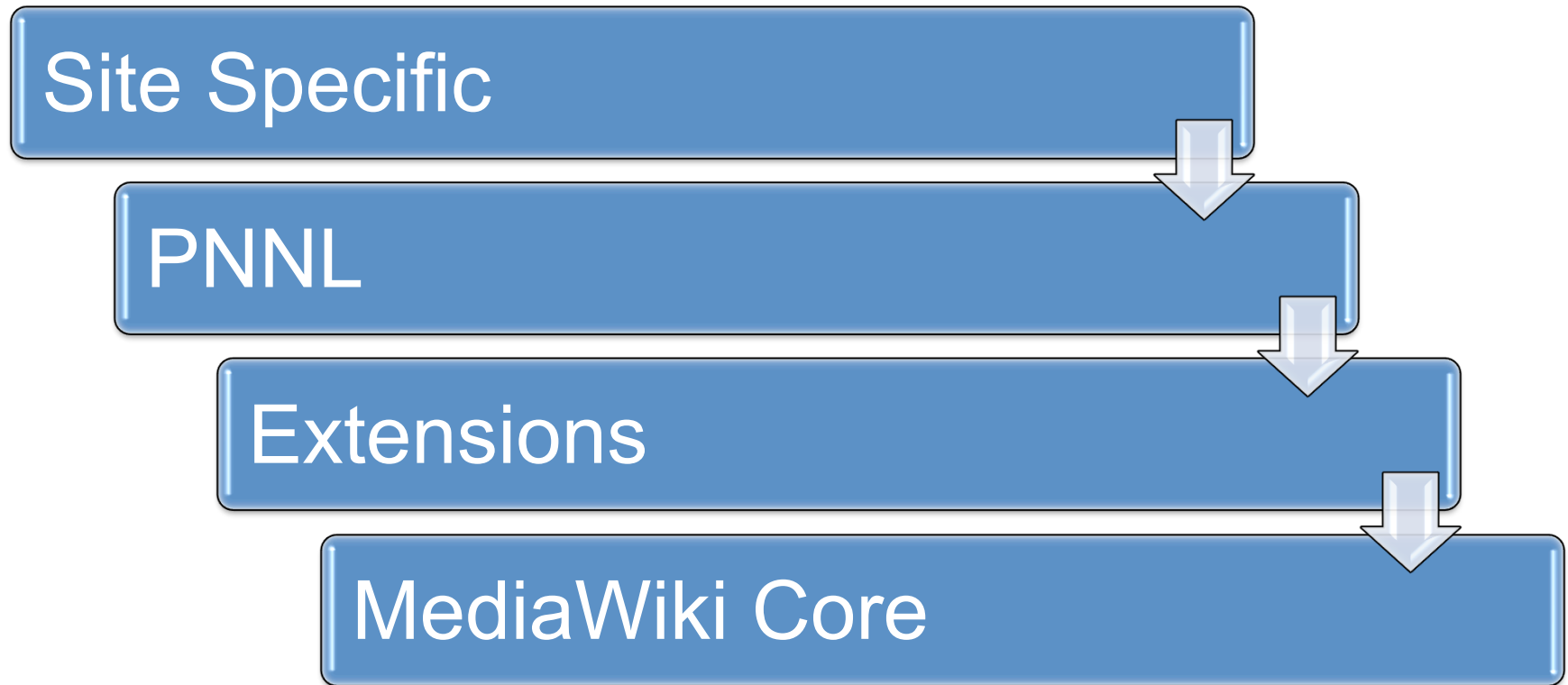
- ▶ Automatic builds when subversion changes to development/test server
- ▶ Manual builds to production or external server
- ▶ Subversion is not cutting it
 - We need separate development and production branches
 - Working on transitioning to git



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

KEF Subversion Setup and Hudson Layering



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

KEF Setup

- ▶ KEF Modules that SMW instances can utilize
 - Automatic property and category creation
 - Automatic image and page creation (templates, Semantic Forms, etc.)
 - XML Specification



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

PNNL Extension



















▼ PNNL	Today, 12:41 PM
▼ includes	Apr 21, 2011 4:37 PM
PNNL_BibTeX.php	Apr 14, 2011 8:27 PM
PNNL_Common.php	Apr 21, 2011 4:37 PM
PNNL_Globals.php	Apr 14, 2011 8:27 PM
PNNL_phpBB_Bridge.php	Apr 14, 2011 8:27 PM
PNNL_Settings.php	Apr 14, 2011 8:27 PM
▼ maintenance	Apr 23, 2011 10:31 AM
categories.xml	Apr 15, 2011 7:08 AM
cleanCategory.php	Apr 14, 2011 8:27 PM
groups.xml	Apr 14, 2011 8:27 PM
KEFCategories.xsd	Apr 14, 2011 8:27 PM
KEFHaloACLGroups.xsd	Apr 14, 2011 8:27 PM
KEFProperties.xsd	Apr 14, 2011 8:27 PM
KEFSemanticData.xsd	Apr 14, 2011 8:27 PM
moveToCategory.php	Apr 14, 2011 8:27 PM
obliteratePages.php	Apr 14, 2011 8:27 PM
PNNL_Setup.inc	Apr 14, 2011 8:27 PM
PNNL_Setup.php	Apr 14, 2011 8:27 PM
properties.xml	Apr 23, 2011 10:31 AM
renameProperty.php	Apr 14, 2011 8:27 PM
▼ modules	Apr 14, 2011 9:36 PM
▶ ADM	Apr 14, 2011 8:29 PM
▶ AnalyticalGaming	Apr 14, 2011 8:29 PM



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Site Specific Overlay

▼  wiki	Today, 12:39 PM
▼  extensions	Today, 12:39 PM
▶  ExhibitExtension	Apr 21, 2011 4:11 PM
▼  PNNL	Today, 12:39 PM
▼  includes	Apr 14, 2011 6:44 PM
 PNNL_Config.inc	Apr 14, 2011 6:44 PM
▼  maintenance	Apr 14, 2011 6:44 PM
 PNNL_Setup.inc	Apr 14, 2011 6:44 PM
▼  modules	Today, 12:43 PM
▶  KEF	Apr 14, 2011 6:44 PM
▼  NSF	Apr 21, 2011 9:35 PM
▼  maintenance	Apr 23, 2011 10:49 AM
▶  images	Apr 14, 2011 8:16 PM
▶  ldrd-data	Apr 23, 2011 10:31 AM
 addpubtemplate.php	Apr 14, 2011 8:13 PM
 importldr.php	Apr 23, 2011 10:49 AM
 metadata.xml	Apr 24, 2011 1:27 PM
 nsfload.php	Apr 14, 2011 8:16 PM



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Example site specific PNNL_Config.inc

```
5 # the PNNL_Setup.php script uses this file to determine what modules are being
6 # included so that they can be initilized.
7 #
8
9 require_once("../modules/Global/maintenance/Global_Setup.inc");
10 require_once("../modules/ITD/maintenance/ITD_Setup.inc");
11 require_once("../modules/BachModel/maintenance/BachModel_Setup.inc");
12 require_once("../modules/KEF/maintenance/KEF_Setup.inc");
13 require_once("../modules/RSS/maintenance/RSS_Setup.inc");
14 require_once("../modules/AnalyticalGaming/maintenance/AG_Setup.inc");
15
16 $setup->complete();
```



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Example site specific PNNL_Settings.inc

```
5 # This file includes those PNNL modules that are required for the IIR site.
6 #
7 global $pnnlDir;
8
9 require_once($pnnlDir . "/modules/BachModel/includes/BachModel_Settings.inc");
10 require_once($pnnlDir . "/modules/KEF/includes/KEF_Settings.inc");
11 require_once($pnnlDir . "/modules/PDF2MediaWiki/includes/pdf2mw_Settings.inc");
12 require_once($pnnlDir . "/modules/AnalyticalGaming/includes/AG_Settings.inc");
13 require_once($pnnlDir . "/modules/Annotator/includes/A7R_Settings.inc");
14 require_once($pnnlDir . "/modules/REST/includes/REST_Settings.inc");
15 require_once($pnnlDir . "/modules/NetLogo/includes/NetLogoHooks.php");
```



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

#5: We needed a means of supporting internal communication and collaboration

- ▶ 15 to 20 sites with several hundred users thus far
- ▶ Internal 'initiative' sites
 - Carbon Sequestration Initiative
 - Microbial Communities Initiative
 - Signature Sciences Initiative
 - And others
- ▶ Project Sites
 - Marine Hydro Kinetic CMS
 - International Building Exchange
 - Naval Postgraduate School LiDAR Research



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

And even more sites

- ▶ Division and Communities of Interest internal sites
 - Computational and Statistical Analytics
 - Visual Analytics, etc.
- ▶ Research
 - National Science Foundation – Science and Innovation Research
 - Automated Content Analysis

#4: We wanted to play with Simile Exhibit.

- ▶ SRF Exhibit didn't do all we needed
- ▶ Discovered a nascent MW Exhibit plugin
 - Added support to do ask queries from the browser
 - Added support for Exhibit collections
 - Added support for Exhibit converters
- ▶ Found disabled extensions and little known, or used facets.



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

More Exhibit goodness...

- ▶ Added new facets
 - 'Saved Filter Facet' saves facet selections as collections

38 TILES • GRAPH • TIMELINE
Items filtered from 1814 originally ([Reset All Filters](#))

sorted by: [labels](#); then by... • ☒ grouped as sorted

Doc020ub.doc.txt (1)

1.

Doc020ub.doc.txt	
Publish Year	2011
Publish Month	2
Publish Day	6
Text	Subhi al-Tufayli, the former

Text Search

☐ **Saved Filters** 2 ☒ [Add](#)

- ☒ All Records ☐
- ☒ aoe ☒
- ☒ oeu ☒

☐ **Promoter**

- 1 Abu-Bakr
- 4 Ahmed Maki

More Exhibit goodness...

► Date selection facet and view

+ Add New Event

TILES • TIMELINE • **CALENDAR** • MAP

5 Events filtered from 17 originally ([Reset All Filters](#))

March, 2011						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
27	28	1	2	3	4	5
					Mike's Newer, Cooler, Event Test	Test2
6	7	8	9	10	11	12
Test3		Test4				

Search

Date

15

☒

Feb March, 2011 Apr

S	M	T	W	T	F	S
27	28	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

2011-03-01 - 2011-03-17

Internal Location

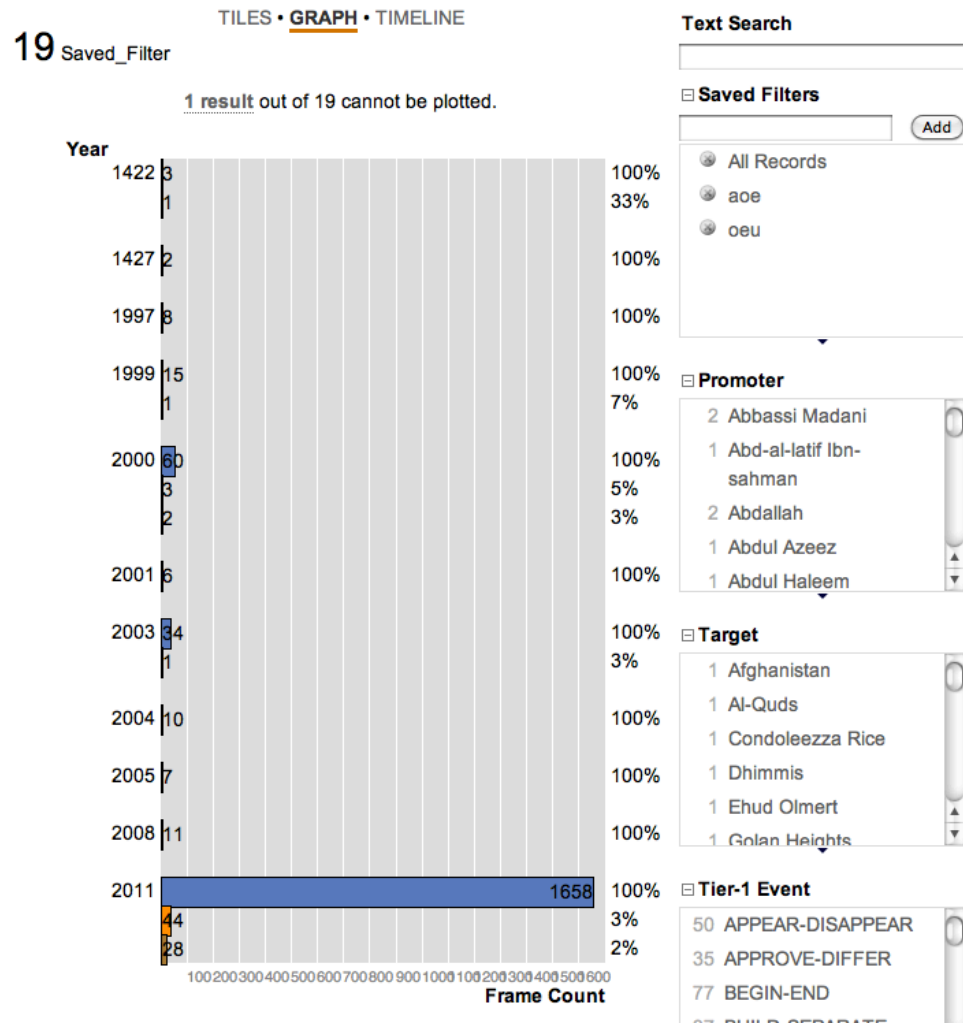


Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

More Exhibit goodness...

► Graph View



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

More Exhibit goodness...

► Property Export Selection View

TILES • FEATURE VECTOR ELEMENTS

▼ General Features

- ☒ Initial Approval Date
- ☒ Duration
- ☒ Fourth Year SW
- ☒ Research Type
- ☒ Number of Awards
- ☒ Sector Code Counts
- ☒ Proposal Ratings

► Publications Features

► Collaboration Features

► IP Features

► Staff Features

► Finance Features



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

More Exhibit goodness...

► Tag Cloud

Cloud

.NET 3D Graphics Agile Development AJAX Algorithms Apache Blog C/C++ C# C++
Cognitive Science Computational Linguistics CSS Database Eclipse Ethnography Flash
Flex Graphic Design HTML Human Computer Interaction Human Factors Human-
Centered Design IIS Information Visualization Java JavaScript Javascript Knowledge
Representation Linux Literature Review Machine Learning MediaWiki MySQL Natural Language
Processing OpenGL Perl Photography Photoshop PHP Program Management Project
Management project management Python RDF Semantic Forms Semantic MediaWiki
Semantic Mediawiki Semantic Web Social Media Analytics Software Architectures SQL SQL
Server Statistics Usability User Experience Design User Interface Design User-Centered
Evaluation Video video and animation Visual Analytics Visual Studio Web Development Web
Services Wiki XML XSLT

267 Staff Members

sorted by: has_name; then by... • ☒ grouped as sorted

1.

Aaron Alva

Phone Number	(509) 375-5930
Email Address	aaron.alva@pnl.gov
Job Title	Intern
Technical Group	Secure Cyber Systems

2.

Aaron Baumann

Phone Number	(509) 372-4549
--------------	----------------

Search

Systems Engineering Integration

1 Intelligence and
Counter Intelligence
Analysis

Advanced Computer Science

1 Data Intensive and High
Performance Computer
Science



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Future Exhibit goodness...

► Radar Graphs



Future Exhibit goodness...

► Relationship Graphs



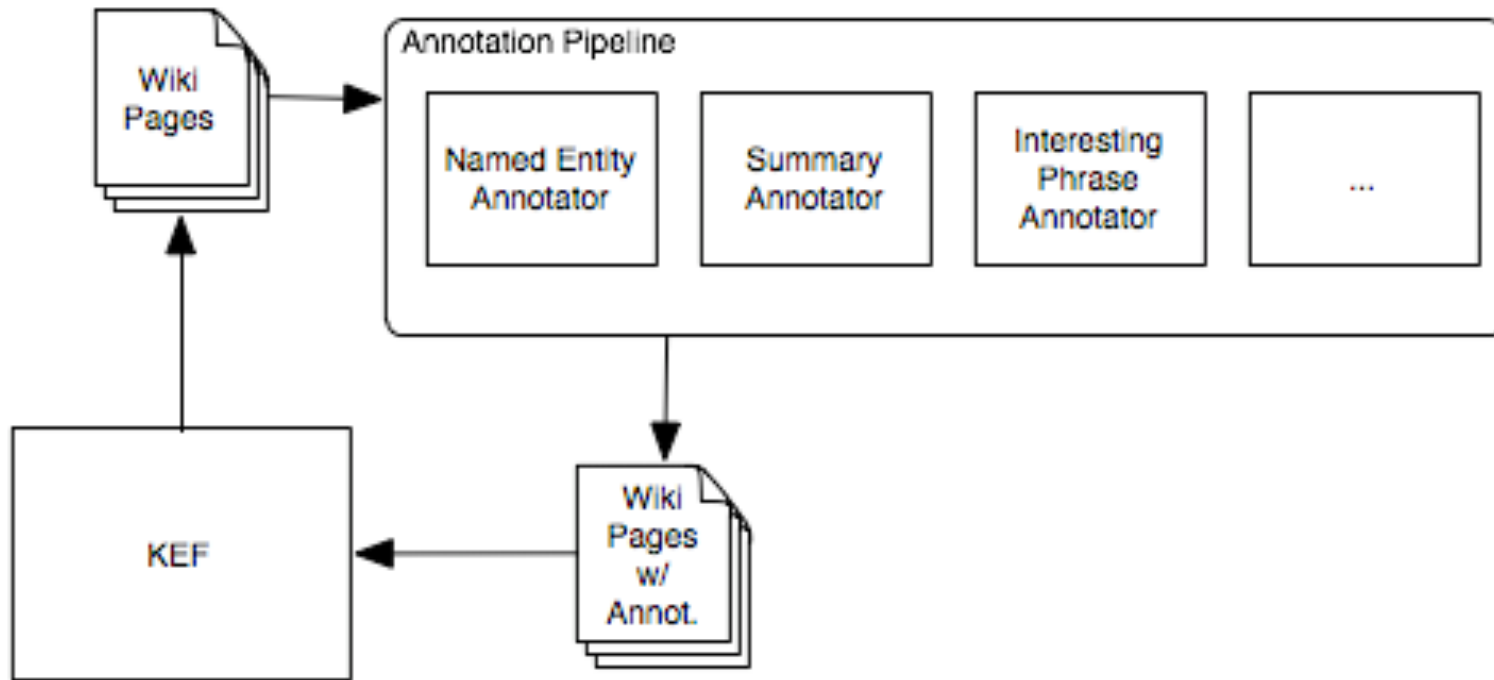
#3: We wanted to build some Natural Language Processing tools.

- ▶ We have developed a Natural Language Processing pipeline that takes unstructured text and adds semantics.
 - People
 - Places
 - Dates
 - Rhetoric
 - Statistically Improbable Phrases
 - Sentiment (target and of groups)
- ▶ Automated Discovery Mechanism finds documents that are similar to 'seed' documents in the wiki.
- ▶ We developed a proof of concept PDF to Wiki markup converter.

Automated Annotation Pipeline (AAP)

1. Pages to be annotated are selected from KEF

2. Pages are processed and annotated by a pipeline of NLP tools



3. The annotated pages are put into KEF as a new revision

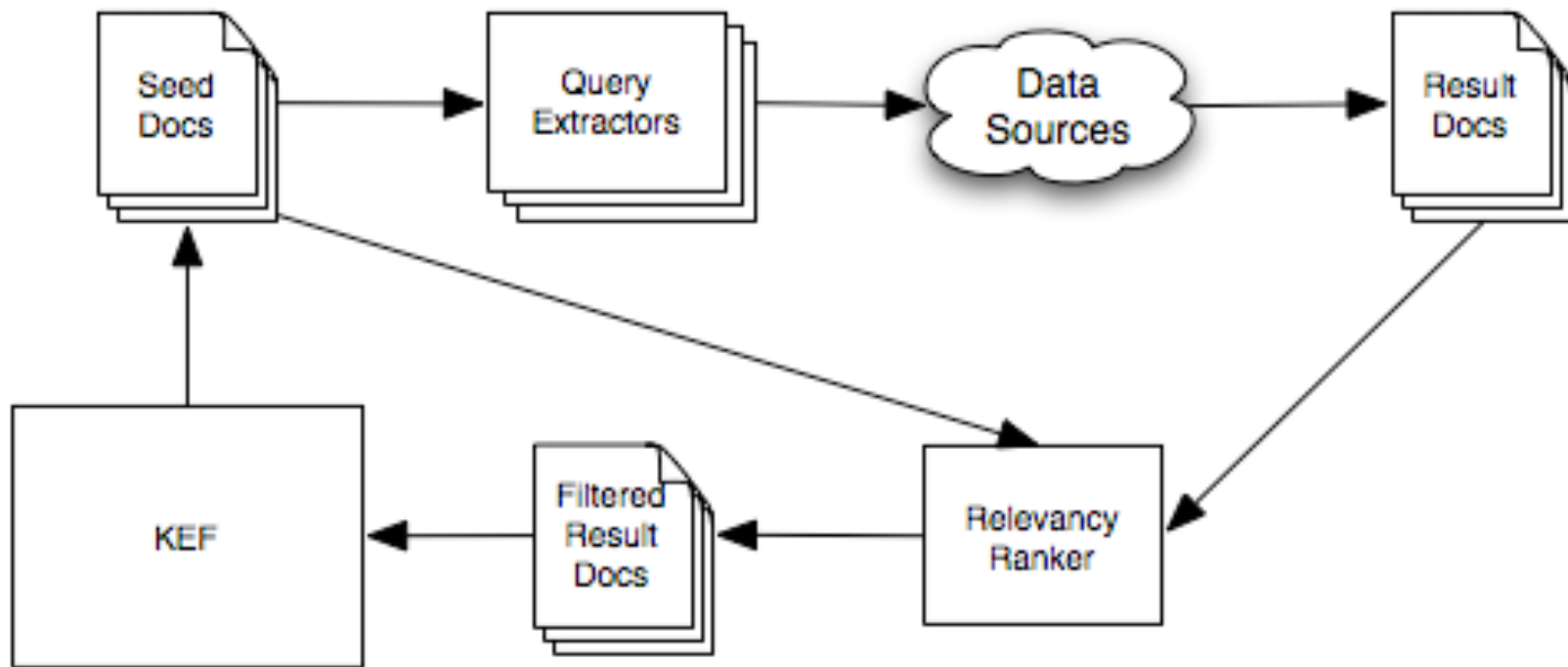


Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Automated Discovery Mechanism (ADM)

1. Seed docs are selected from KEF
2. Queries are extracted from the seed docs
3. Search terms are sent to data sources and results are harvested



5. The most relevant result docs are put in KEF

4. Relevancy of each result is calculated



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

#2: We wanted to write a Wiki markup parser.

- ▶ One of the KEF requirements was an GUI annotation tool that could
 - Nest annotations from the NLP pipeline
 - Create annotations across paragraph boundaries.
- ▶ Started by hacking on the Halo Advanced Annotation Tool. But I just couldn't get something working that was clean.
- ▶ So I build a real wiki text parser that generated AST's that I then used to render the wiki text in a Sproutcore application.



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Parser details

- ▶ PEG parser generator to deal with the vagaries of wiki text like the quote mess
- ▶ Not fully implemented, e.g., tables and images don't work
- ▶ Generates JavaScript, so it runs in the browser.
- ▶ It is SLOW. Firefox and MSIE are nearly unusable, but Safari works well.
- ▶ A new version will likely create the AST's on the server and cache them.



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Annotation Details

- ▶ Store annotations as Semantic Internal Objects, with properties for
 - Property name(s)
 - Property value(s)
 - Text
 - Text offsets



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Obligatory Screen Shot

personnel or hardware inside nuclear facilities, protection for classified information, and no compromise of domestic nuclear security and safeguards. Additional testing used effectively in future arms control

Introduction
At the **Helsinki** summit of **March 21** **Strategic Arms Reduction Treaty**, relating to the reduction of nuclear

A bilateral agreement on nuclear w **Russian** national security.² reductions far more difficult to reveal could be re-deployed in active stockpiles. reductions in nuclear arsenals unless warhead and fissile material treaty

To provide these potential benefits warheads must overcome a range of challenges. be hidden or diverted from a storage facility. the number of nuclear warheads that could be lost. loss of classified information while in compliance.⁴ The nuclear arms control treaty for a warhead arms control treaty would take place are some of the most sensitive government installations in the **United States** and

es that such systems could be

framework agreement for a third s control measures directly so poses significant challenges.

potential benefits for **U.S.** and could make nuclear arms near warheads, these weapons l prevent much deeper materials are established. A and secure.

e dismantlement of nuclear ads are relatively small and can **States** and **Russia** keep secret need to be taken to prevent the m treaty assembly plants where inspections

for a warhead arms control treaty would take place are some of the most sensitive government installations in the **United States** and

Case Study: Nigeria Intent to Proliferate

Indicator: Select an Indicator

Event: Select an Event

Support: Support Refute

Credibility / Bias

Cancel Save Changes



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

#1: We wanted to work on an open source project.

- ▶ We are working on being able to contribute to MediaWiki and Semantic MediaWiki and other extensions
 - Bug fixes
 - Enhancements
- ▶ We are in the planning stages of open sourcing KEF.
 - Currently specifying what will be included
 - NLP pipeline (parts of it)
 - Annotator
 - Build system
 - What else?



Pacific Northwest
NATIONAL LABORATORY

The end...

Thanks!